元数据管理平台
技术白皮书

北京亿信华辰软件责任有限公司
2018 年 4 月
目录

1. 前言 .................................................................................................................. 1
   1.1. 关于本白皮书 ......................................................................................... 1
   1.2. 背景介绍 ................................................................................................. 1
   1.3. 产品定位 ................................................................................................. 1

2. 产品架构 ........................................................................................................ 2
   2.1. 概述 ........................................................................................................ 2
   2.2. 数据源层 ................................................................................................. 2
   2.3. 采集层 .................................................................................................... 2
   2.4. 数据层 .................................................................................................... 3
   2.5. 功能层 .................................................................................................... 3
   2.6. 访问层 .................................................................................................... 3

3. 产品功能特色 ................................................................................................ 4
   3.1. 规范的元模型管理 ................................................................................ 4
   3.2. 端到端的自动化采集 ............................................................................ 5
   3.3. 全面的采集适配器 ................................................................................ 5
   3.4. 可灵活定制的采集模板 ....................................................................... 6
   3.5. 便捷的元数据检索 ............................................................................... 7
   3.6. 完善的元数据管理 ............................................................................... 7
   3.7. 强大的元数据版本管理 ....................................................................... 8
   3.8. 实时的元数据变更监控 ....................................................................... 8
   3.9. 数据地图鸟瞰全局 ............................................................................... 9
   3.10. 丰富的元数据分析应用 ................................................................. 9
       3.10.1. 血缘分析 ...................................................................................... 9
       3.10.2. 影响分析 ..................................................................................... 10
       3.10.3. 全链分析 ..................................................................................... 10
       3.10.4. 关联度分析 ............................................................................... 11
       3.10.5. 属性差异分析 ........................................................................... 11
   3.11. 出色的元数据检核机制 ................................................................. 12
       3.11.1. 一致性检核 ............................................................................... 12
       3.11.2. 属性填充率检核 ....................................................................... 12
       3.11.3. 组合关系检核 ........................................................................... 12
   3.12. 自助式门户 ....................................................................................... 13
   3.13. 丰富服务接口 ..................................................................................... 13

4. 产品技术优势 ................................................................................................ 13
4.1. 系统设计原则 ........................................................................................................... 13
  4.1.1. 先进性 ........................................................................................................... 14
  4.1.2. 可维护性 ...................................................................................................... 14
  4.1.3. 可靠性 .......................................................................................................... 14
  4.1.4. 易用性 .......................................................................................................... 15
  4.1.5. 安全性 .......................................................................................................... 15
  4.1.6. 扩展性 .......................................................................................................... 15
4.2. 可扩展采集适配器设计 ...................................................................................... 16
4.3. 采用 MOF 规范 .................................................................................................. 16
4.4. 支持基于 XMI 的数据交换 ............................................................................... 17
4.5. 运用 RESTful 架构 ............................................................................................ 18

5. 软硬软件环境 ......................................................................................................... 19
  5.1. 服务器配置推荐 ............................................................................................... 19
  5.2. 客户端配置 ....................................................................................................... 20
    5.2.1. 客户端(建议配置) .................................................................................... 20
    5.2.2. 客户端浏览器 .......................................................................................... 20
1. 前言

1.1. 关于本白皮书

本白皮书对应产品版本为：元数据管理技术白皮书 V2.1。

最后修订日期：2018 年 05 月。

本白皮书将在阐述元数据产品的背景介绍之后，详细介绍元数据管理平台在技术和功能上的特点。

1.2. 背景介绍

随着数字化加速，数据量呈指数增长，大数据相关技术的出现，让大家看到了新机遇的同时，大家对数据治理的需求也在增加。尤其是大型企业业务数据、政府政务数据、行业标准数据，往往由于数据分散、质量参差不齐、数据存储结构差异大，虽然数据中蕴含大量有价值信息，但想要挖掘出来，往往需要做大量的数据治理工作，成本十分高昂。Gartner 分析师在 2017 年预测，到 2020 年，50%的信息治理举措将通过基于元数据的政策制定。就国内而言，目前完备成熟的元数据产品较少，市场正处于上升期，是抢占市场份额的好时机。

1.3. 产品定位

亿信元数据管理平台致力于处理技术元数据、业务元数据、管理元数据，帮助各行各业用户获得更多的数据洞察力，进而挖掘出隐藏在资源中的价值。

对技术人员而言，元数据管理平台通过对将分散、存储结构差异大的资源信息进行描述、定位、检索、评估、分析，实现了信息的描述和分类的结构化，从而为机器处理创造了可能，大大降低数据治理人工成本。正因如此，元数据已经成为了很多大型数据治理项目的核心。

对业务人员而言，元数据管理平台通过对业务指标、业务术语、业务规则、业务含义等业务信息进行描述、定位、检索、评估、分析，协助业务人员理解业务含义、
行业术语和规则，业务指标取数据口径和影响范围等。

2. 产品架构

2.1. 概述

元数据管理平台架构分为 5 层，数据源层、采集层、数据层、功能层和访问层，下图为元数据管理产品的整体架构图。

数据源层是指元数据管理平台所支持的元数据来源的方式。提供直连多种不同类型的数据源，包括：数据库类型、ETL 类型、文件类型、业务系统类型等。

2.3. 采集层

采集层针对不同数据源提供丰富的适配器，实现端到端的自动化采集。具体包括：sqlserver、oracle、mysql、postgresql、petabase、ODI、Excel、亿信 BI 等。

同时支持适配器扩展，实现最大限度的自动化采集。
2.4. 数据层

元数据数据层是基于关系数据库的元数据存储，用于实现元数据和元模型的数据的物理存储。元模型存储元数据的属性要求和存储格式要求。元数据存储了从各个系统中采集而来的元数据信息。

2.5. 功能层

元数据功能层提供了元数据管理产品的基本功能，包括元模型增删查改及版本发布功能、元数据增删查改及版本管理、元数据变更管理、元数据分析应用、元数据检核以及产品的系统管理功能。

其中元模型管理模块用于操作元模型，元模型是对各个种类元数据以及元数据之间关系的定义，元模型包括两部分：一部分由元数据管理平台产品内置的标准元模型，另一部分是用户根据管理需求自定义的元模型。元模型管理还设计了发布功能，只有在发布之后才会生效，使用户在设计元模型时，不会影响到元数据的使用。

元数据管理主要包括了元数据增删改查日常维护，版本管理，元数据全文检索。

元数据分析应用主要包括了血缘分析、影响分析、关联度分析、数据地图等多种图形化分析应用，并提供导出和收藏功能，将分享结果进行留档。

元数据检核包括一致性检核、属性填充率检核和组合关系检核，是保障元数据质量的重要手段之一

系统管理功能包括了机构用户角色的权限管理、系统备份恢复、门户应用、日志管理、系统监控等系统运维相关的功能。

2.6. 访问层

元数据访问层用于给用户提供访问控制服务。元数据产品面向的主要用户群有三类：技术设计人员、业务分析人员、以及系统的运维人员。通过门户访问和后台访问，可以实现多种角色的访问控制。同时访问层还提供了多种形式的接口服务，可以很方便的与其它 IT 系统进行集成。

ESENSOFT
3. 产品功能特色

3.1. 规范的元模型管理

亿信元数据管理平台元模型以Meta Object Facility (MOF)规范为基础，支持XMI格式的元模型导入和导出，同时内置大量技术元数据、业务元数据的元模型，用户可直接使用。元模型管理对元模型的基本信息、属性、父子关系、依赖关系、组合关系的增删改查操作，内置元模型的内置信息不允许修改或者删除，但可进行新增操作。

具体功能界面如下：

![元模型查看界面](image1)

图2 元模型查看界面

元模型支持发布功能，只有发布后的元模型才可被元数据使用，同时支持查看所有发布版本。

![元模型版本查看界面](image2)

图3 元模型版本查看界面

通过发布过程，将元模型的设计和运用隔离开，元模型只有在发布之后才会生效，
使用户在设计完成发布之前，不会影响到元数据的使用。

3.2. 端到端的自动化采集

对元数据信息的维护除界面手动操作方式外，亿信元数据管理平台利用内置采集适配器，让用户通过配置数据源参数及定时采集任务，进行自动化采集。实现直连数据源的端到端元数据采集。

![采集任务配置界面](image)

图4 采集任务配置界面

3.3. 全面的采集适配器

元数据管理系统提供了丰富的内置适配器，来保证自动化采集的同时，还支持对适配器进行扩展。

本版本内置适配器：

- 亿信 BI 采集适配器
- i@Report 采集适配器
- JDBC 驱动采集驱动适配器
- Greenplum 采集适配器
- MySOL 采集适配器
- Oracle 采集适配器
3.4. 可灵活定制的采集模板

采集模板定制允许用户根据自己的采集需求，灵活选择需要采集的元数据及其属性，并自动生成 EXCEL 采集适配器可识别的 EXCEL 模板文件。用户可使用模板文件将数据批量录入。

图 5 采集模板配置界面
3.5. 便捷的元数据检索

提供对元数据的全文检索功能。检索支持对检索范围、检索类型、修改时间进行过滤，过滤条件支持保存，让用户可以将常用的过滤条件保存使用，以便能够更加快速浏览所需元数据。

图6 元数据检索界面

3.6. 完善的元数据管理

元数据管理功能提供各类元数据管理，包括：业务元数据、技术元数据和管理元数据，支持元数据的基本信息、属性、依赖关系、组合关系的增删改查操作。

图7 元数据管理界面
3.7. 强大的元数据版本管理

元数据管理平台提供元数据版本管理功能，可以对元数据进行发布、查看历史版本、导出历史版本、版本对比操作。

图8 版本发布界面

在元数据没有发布时，仅有使用权限的用户不会获得未定版信息，保证了用户使用元数据系统的权威性和可靠性。

3.8. 实时的元数据变更监控

元数据管理平台可实时对元数据变更进行监控，并提供变更订阅功能，将用户关心的元数据的变更情况定期发送到用户邮箱。

图9 变更查询界面
3.9. 数据地图鸟瞰全局

一般情况下，元数据管理的业务繁多、形式各异，在集中管理后，如何提供便捷的使用方式，是发挥信息资产价值的关键。数据地图从宏观层面组织信息，力求以用户视角对信息资产进行归并、整理，全局展现资产的宏观信息，有效挖掘信息的潜在价值。

![数据地图](image)

图10 数据地图

3.10.丰富的元数据分析应用

元数据管理平台提供了丰富的分析应用，包括：血缘分析、影响分析、全链分析、关联度分析、属性差异分析，同时支持将分析结果进行导出和收藏。

3.10.1. 血缘分析

血缘分析是对指定元数据的起源及其推移位置的分析。它反应数据的来源与加工过程，还描述了数据在不同过程中发生的情况。它可以帮助分析信息的使用方式并追踪用于特定用途的关键信息位。
3.10.2. 影响分析

影响分析帮助用户迅速了解分析对象的下游数据信息，快速掌握元数据变更可能造成的影响，以便更有效的评估变化该元数据带来的风险，从而帮助用户高效准确的对数据资产进行清理、维护与使用。

3.10.3. 全链分析

全链分析是用来分析指定元数据前后与其有关系的所有元数据，不仅反应了元数据的来源与加工过程，也反应了元数据的使用情况，使用全链分析可清晰的了解该元
数据的来龙去脉。

### 3.10.4. 关联度分析

关联度分析是从关系数量的角度对指定元数据进行分析，来体现该元数据在系统中依赖程度的高低，从一定的角度可以反映出该元数据的重要程度。

### 3.10.5. 属性差异分析

属性差异分析是用来比较同类型元数据之间属性值的差异，方便用户识别相似元
数据之间的存在的微小差距。

<table>
<thead>
<tr>
<th>元数据路径</th>
<th>A2</th>
<th>B2</th>
<th>C2</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>元数据路径</td>
<td>A2</td>
<td>B2</td>
<td>C2</td>
<td>D2</td>
</tr>
<tr>
<td>元数据路径</td>
<td>A2</td>
<td>B2</td>
<td>C2</td>
<td>D2</td>
</tr>
<tr>
<td>数据模型表</td>
<td>A2</td>
<td>B2</td>
<td>C2</td>
<td>D2</td>
</tr>
</tbody>
</table>

图15 属性差异分析

3.11.出色的元数据检核机制

由于元数据是很多数据管理活动的基础，所以所有类型的数据中，元数据的质量是最为重要的。亿信元数据管理平台提供元数据质量检核功能，包括一致性检核、属性填充率检核和组合关系检核，是保障元数据质量的重要手段之一。

3.11.1 一致性检核

一致性检核用来检验来源系统元数据以及元数据之间信息是否一致的功能，帮助管理人员分析出元数据管理平台与来源系统之间的差异，辅助管理人员更好的维护元数据管理平台。

3.11.2 属性填充率检核

属性填充率检核是用来检验选定范围内元数据属性的填充情况，方便用户了解某些重要元数据的属性的填充情况。

3.11.3 组合关系检核

组合关系检核用根据元模型中设置的组合和被组合关系对元数据进行检查，查找出不符合元模型定义的组合关系的元数据，例如找出没有与任何数据库表建立组合关系的字段。
3.12.自助式门户

元数据管理系统带有门户工具是元数据，是实现个性化、分业务群的元数据及相关应用展现的重要手段，为用户关心的内容提供快速访问的通道，节省时间和提供新知识。

用户只需通过简单的拖拉控件到定义门户区域，就能自己定义个性化的门户，能够简单实现各种复杂的自定义用户视图，无需开发人员以及美工介入。

图16 门户展现

3.13.丰富的服务接口

元数据管理平台，提供丰富的接口以支持与外部系统进行交互。

- 支持嵌入其它平台
- 支持通过接口进行元数据信息查询
- 支持通过接口获取元数据变更信息
- 提供接口支持对适配器进行扩展

4. 产品技术优势

4.1. 系统设计原则

元数据管理平台是基于 B/S 架构的软件平台，研发过程中遵循软件工程迭代开发
和技术白皮书

和面向对象设计思想。运用了先进的软件开发技术，元数据管理平台的设计遵循下列原则：

### 4.1.1. 先进性

- SQL 解析技术
- 全文检索技术
- 采用 Meta Object Facility (MOF) 规范
- 支持 XMI 数据交换标准
- 采用 RESTful 架构
- 支持标准的 OLAP 分析
- 遵循事实结构的星型数据仓库模型
- 用户 WEB 界面采用 AJAX 技术
- 使用国际化编码 UTF-8

### 4.1.2. 可维护性

- 提供集成化的系统管理模块，通过图形化界面管理、配置所有系统对象
- 元模型可方便的进行备份和恢复
- 支持自定义元模型
- 纯 web 化，无需任何插件，客户端零安装，大大提高系统升级维护
- 所有系统元数据全部持久保存在数据库中，服务器迁移十分方便

### 4.1.3. 可靠性

- 合理的 JAVA 内存回收机制
- 有完善的错误诊断和恢复机制，不会出现因用户误操作而导致系统崩溃
- 所有系统元数据全部持久保存在数据库中，不会因应用服务器崩溃或病毒入侵等损坏系统

ESENSOFT
4.1.4. 易用性

- 系统安装配置简单，基于 B/S 结构，客户端零安装
- 界面操作简单，流程清晰
- 面向业务人员提供零编程、交互式元模型设计功能
- 广泛采用 AJAX 技术，为用户提供了十分友好的交互式 WEB 操作界面

4.1.5. 安全性

- 通过元模型的版本管理，保障元数据在使用时的稳定性
- 管理员可详细配置每个用户的操作权限和可访问对象
- 采用严格的技术手段，杜绝非法用户绕过安全认证直接获取元数据信息
- 记录详细的操作和采集日志以备审查
- 所有密码均使用安全的加密算法加密保存
- 防暴力攻击
- 防 SQL 注入攻击，跨站脚本攻击等
- 支持 SSL 通信协议

4.1.6. 扩展性

- 采集适配器采用可扩展设计，用户可自行上传针对自己某个业务系统的适配器
- 所有功能都是根据可扩展可通用的原则来设计的，可以适应未来未知的需求变化
- 系统的模块化设计，一个模块发生变化不会影响整个系统
- 提供各种定制接口和调用接口 Web Services 服务，满足第三方服务需要
- 提供二次开发平台设计，满足未来各种未知需求
4.2. 可扩展采集适配器设计

元数据管理平台提供完备的采集适配器设计接口，并采用 Java 动态加载技术，实现 jar 包的动态加载，使用户在无需重启服务器的情况下，动态将自己设计的适配器加载到元数据管理平台中，以实现个性化数据的自动化采集，达到最大限度的自动化采集。

4.3. 采用 MOF 规范

元数据管理平台采用 Meta Object Facility (MOF) 规范作为元数据管理的基础。

MOF 称为元对象设施或者元对象机制，是 OMG (Object Management Group，对象管理组织) 为了帮助销售商、开发者和用户更好的使用元模型和元数据技术而制定的。MOF 最初是 OMG 在 1996 年一个征求建议书 (RFP, Request For Proposal) 中提出的。2002 年 MOF 成为 OMG 建模系列规范中的一员，2005 年 MOF 规范成为 ISO (ISO/IEC 19502) 的标准。因此，MOF 也可以称为 MOF 规范或者 MOF 标准。符合 MOF 规范的元模型具有开放性、扩展性和互操作性。

MOF 规范对应着一个 MOF model, 它实际上也是一个元元模型 (meta meta-model)，它定义了描述元模型的描述语言并用这个语言描述了自身。虽然 MOF model 是一个强大的描述大量的信息模型的建模语言，但是它不是最终的建模语言，更多的使用 MOF 是把它作为一个规范和工具，去设计和实现更优秀的元模型建模系统。

MOF 规范位于模型体系最底层，所有模型体系规范的基础都是 MOF 规范，包括 UML，CWM 都是由 MOF 扩展而来。

基于 MOF 的还有模型交换的规范 XMI, 为不同元数据交换提供了很好的模型基础。
若整个元数据围绕 MOF 设计和扩展，不用修改元数据管理核心部分，就可以适应元数据种类的不断扩展。

一般将元数据有关的体系分三层，M0(数据)、M1(元数据)、M2(元模型)、M3(元元模型)，其中 MOF 元元模型中描述了包、元素、属性、命名空间和约束等对象及其关系，位于层次结构的最上层，也是最抽象的一层。

图 18 元数据体系

以 MOF 作为底层元元模型来支持元数据管理，在 M2 层中就可以对元模型进行定义和扩展（例如 CWM 模型），将来还可以扩展到微服务模型、业务模型等。

4.4. 支持基于 XMI 的数据交换

XMI (XML-based Metadata Interchange) 是基于 XML 的元数据交换。它通过标准化的 XML 文档格式和 DTDs (Document Type Definitions) 为 UML 元模型（元模型是一类特殊的模型）和其他模型定义了一种基于 XML 的数据交换格式。它同时也定义了一个从 UML 到 XML 的映射。

XMI 的主要目的就是让各种分布式的异构环境中的建模工具和元数据存储 (metadata repositorie) 仓库之间能方便地进行数据交换。

XMI 规范中包括了两个主要的部分:
XML DTD 的产生规则：用于为使用 XMI 进行编码的元数据文件产生相应的 XML DTD 文件。XMI DTDs 作为 XMI 文档的语法描述文件，可以方便地使用通用的 XML 工具对 XMI 文档进行有效性校验。

XML 文档的产生规则：它为元数据到纯 XML 文档提供了编码规则。同样它还应用于把 XMI 文档反编码，重新构造生成元数据。

XMI 规范支持任何可以用 MOF 表示的元数据（包括模型和元模型）的数据转换。规范同时支持完整的模型或是一个模型的片断到 XML 的转换。

元数据管理平台支持 XMI 格式的元模型导入和导出。

4.5. 运用 RESTful 架构

RESTful（Representational State Transfer，表现层状态转化）是 Roy Thomas Fielding 在 2000 年他的博士论文《Architectural Styles and the Design of Network-based Software Architectures》中提出的一个描述互联系统架构风格的名词。让我们先去理解 Representational State Transfer 这个词组到底是什么意思？Web 本质上由各种各样的资源组成，资源由 URI 唯一标识。浏览器将展示出该资源的一种表现方式，或者一种表现状态。如果用户在该页面中定向到指向其它资源的链接，则将访问该资源，并表现出它的状态。“表现层”其实指的是“资源”的“表现层”，这意味着客户端应用程序随着每个资源表现状态的不同而发生状态转移，也即所谓 RESTful。

RESTful 架构用基于标准的 HTTP(s) 协议实现资源的增/删/改/查 (CRUD) 操作：

- **GET**：类似 SQL 的 SELECT 语句，从服务器取出资源（一项或多项）。
- **POST**：类似 SQL 的 INSERT 语句，在服务器新建一个资源。
- **PUT**：类似 SQL 的 UPDATE 语句，在服务器更新资源（客户端提供改变后的完整资源）。
- **PATCH**：类似 SQL 的 UPDATE 语句，在服务器更新资源（客户端提供要改变的属性）。
- **DELETE**：类似 SQL 的 DELETE 语句，从服务器删除资源。
元数据管理平台遵循 RESTful 规范，提供 RESTful API 接口为第三方应用提供服务。

5. 软硬软件环境

5.1. 服务器配置推荐

<table>
<thead>
<tr>
<th>配置项目</th>
<th>说明</th>
</tr>
</thead>
<tbody>
<tr>
<td>中间件</td>
<td>IBM X3650 或 HP PowerEdge R910</td>
</tr>
<tr>
<td></td>
<td>CPU：Xeon E7-4820</td>
</tr>
<tr>
<td></td>
<td>内存：32G</td>
</tr>
<tr>
<td></td>
<td>硬盘：RAID1 1T</td>
</tr>
<tr>
<td></td>
<td>网络：1000M</td>
</tr>
<tr>
<td>数据库配置</td>
<td>IBM X3850 或 HP PowerEdge R720</td>
</tr>
<tr>
<td></td>
<td>CPU：Xeon E7-4820*2</td>
</tr>
<tr>
<td></td>
<td>内存：128G</td>
</tr>
<tr>
<td></td>
<td>硬盘：按需</td>
</tr>
<tr>
<td></td>
<td>网络：1000M</td>
</tr>
<tr>
<td>操作系统及软件</td>
<td>操作系统：Windows、linux</td>
</tr>
<tr>
<td></td>
<td>中间件：tomcat-7.0.70</td>
</tr>
<tr>
<td></td>
<td>数据库：Oracle11g、Oracle12c</td>
</tr>
</tbody>
</table>

说明：

若采集数据大于 20W 条，建议数据库内存 6~8G，redo log 分 3 组，每组 10G，undo 回滚空间至少分配 30G，用户表空间建议初始化为 30G。
5.2. 客户端配置

5.2.1. 客户端（建议配置）

<table>
<thead>
<tr>
<th>所属项</th>
<th>说明</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>不低于酷睿 i3 2350M</td>
</tr>
<tr>
<td>内存</td>
<td>不小于 4G</td>
</tr>
<tr>
<td>网络</td>
<td>不低于 100KB/s 传输速度</td>
</tr>
</tbody>
</table>

5.2.2. 客户端浏览器

<table>
<thead>
<tr>
<th>操作系统</th>
<th>浏览器</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows</td>
<td>IE11,IE10,IE9,Chrome 浏览器</td>
</tr>
</tbody>
</table>